STAT 621 Chapter 5

Zepu Zhang
November 10, 2011

General comments:

1. Suppose two independent samples, $\{X_i\}_1^n$ and $\{Y_i\}_1^m$, are drawn from the same distribution, then if we put the two samples together and find the ranks, (then the X ranks have equal chance to be any n numbers in $1, \ldots, N$, where $N = n + m$. This is starting point of analysis in rank-based methods.
2. Deal with ties: average ranks.
3. The `R` function `rank` returns ranks and takes care of ties this expected way.
4. Exact analysis in this chapter all require the pooled sample to have no ties or very few ties. When ties are plenty, use approximate distributions (standard normal obtained from standardization).

Useful results from Chapter 2:

1. $$1+2+\cdots+n = \frac{n(n+1)}{2}, \quad 1^2+2^2+\cdots+n^2 = \frac{n(n+1)(2n+1)}{6}$$
2. Randomly select n numbers from $\{1, 2, \ldots, N\}$, $n \leq N$, without replacement. Let the selected numbers be $X_1, \ldots, X_n$ (they are all random variables) and denote $Y = \sum_{i=1}^n X_n$. Then

$$E(X_i) = \frac{N+1}{2}, \quad \text{var}(X_i) = \frac{(N+1)(N-1)}{12}$$

$$E(Y) = \frac{n(N+1)}{2}, \quad \text{var}(Y) = \frac{n(N+1)(N-n)}{12}$$

1 Rank test for equal mean of two independent samples

Reasons for using ranks instead of the actual data: (1) In some cases the numbers assigned merely to indicate orders (the variable is on ordinal scale); the numbers do not have more numerical meanings; (2) If the distribution is non-normal, exact null distribution is usually very difficult to find whereas statistics for ranks is easier and often does not depend on the actual distribution; (3) Using ranks does not lose much efficiency or power compared to parametric methods that rely on distributional assumptions.

The method is known as the Mann-Whitney, or Wilcoxon, test. It tests that two independent samples come from the same distribution:

H_0: $F(x) = G(x)$ for all x
H_a: $F(x) \neq G(x)$ for some x

Note On notation: F and G denote the CDF of X and Y, respectively. We are not using $G(y)$ because we want to say F and G are equal at the same input value, say x. x here is just a symbol for a number.

Note Alternatives: (1) Alternative hypothesis may be two-sided or one-sided. A one-sided alternative looks like $F(x) > G(x)$ or $F(x) < G(x)$. (2) The test can also be used to test for equal means.

Test statistic is the total rank of X:

$$T = \sum_{i=1}^{n} R(X_i)$$

1.1 Null distribution when there are no ties

There are $\binom{n+m}{n}$ ways to assign ranks to $\{X_i\}$, each with probability $1/\binom{n+m}{n}$. Use computer, we can list all these rank assignments, calculate T in each case (some values may be shared by multiple rank assignments), and make a table for the distribution distribution of T.

Based on this exact distribution, we can calculate the (two-sided) P-value.

We know $E(T) = \frac{n(N+1)}{2}$ and $\text{var}(T) = \frac{nm(N+1)}{12}$

In large-sample cases, we can normalize T using this info to get an approx standard normal test statistic. We can also use this approx normality to calculate the P-value.

1.2 Null distribution when there are (many) ties

In this case the mean stays at $n(N+1)/2$, but the variance of T is (approximately, I believe)

$$\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}$$

Use them to standardize T and get a standard normal test statistic. If we calculate this variance in the case of no ties, it turns out to be $nm(N+1)/12$. Hence it is not necessary to ponder whether the number of ties is large or not; simply use the modified form for variance.

Example EX 1, page 276.

Example EX 2, page 278.

2 Squared rank test for equal variances

Because $\text{var}(X) = E[(X-\mu_X)^2]$, the test is whether the samples $\{(X_i-\mu_X)^2\}$ and $\{Y_j-\mu_Y)^2\}$ have the same mean. We'll use the idea of rank-test for equal mean.

Define

$$U_i = |X_i - \mu_X|,\ i = 1, \ldots, n$$
$$V_j = |Y_j - \mu_Y|,\ j = 1, \ldots, m$$

Pool the two samples and get the ranks. Note that we should really use $(X_i-\mu_X)^2$ and $(Y_j-\mu_Y)^2$ but the ranks are not affected by the squaring. Usually μ_X and μ_Y are replaced by sample means because population means are unknown.

Test statistic:

$$T = \sum_{i=1}^{n} [R(U_i)]^2$$

Remark Why do we use squared ranks instead of ranks? Notice that the test is about $(U_i-\mu_X)^2$ and $(V_j-\mu_Y)^2$, but their ranks do not reflect any difference made by the squares. Informally thinking, somehow the squared ranks get the scale right.

2.1 Null distribution when there are no ties

The exact distribution of T can be found by exhaustively listing all the possible rank assignments and their corresponding T values, just like what is done for the rank test for equal means.

The mean and variance of T are

$$E(T) = \frac{n(N+1)(2N+1)}{6}, \quad \text{var}(T) = \frac{mn(N+1)(2N+1)(8N+11)}{180}$$

In large-sample situations, these can be used to get a standard normal test statistic, or to calculate the P-value.

2.2 Null distribution when there are (many) ties

In this case the mean and variance of T are modified to be (approximately, I believe)

$$E(T) = n\overline{R^2}, \quad \text{var}(T) = \frac{nm}{N(N-1)} \sum_{i=1}^{N} R_i^4 - \frac{nm}{N-1} \left(\overline{R^2}\right)^2$$

Use them to standardize T and get a standard normal test statistic for conducting the test and calculating the P-value.

Exercise Check that these mean and variance, if used without ties, are equal to the no-tie versions. (The mean is easy to verify using analytical results. The variance may be checked numerically.)

3 Measures of rank correlation

Traditional requirements of a correlation measure: p 312.

Pearson's product moment correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\left(\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2\right)^{1/2}\left(\sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2\right)^{1/2}}$$

Note

1. r uses the actual data values rather than ranks.
2. r reflects linear association only.
3. The null distribution of r (i.e. when $r = 0$) depends on the actual joint distribution of (X, Y). This is often unknown, hence r is not useful in tests.

We'll introduce two measures of correlation based on ranks: rank X_i among X, and rank Y_i among Y. Do not mix X and Y; for the purpose of correlation one does not need to "cross-compare" X and Y.

When we conduct tests about the association between X and Y using these measures, the null hypothesis is not that X and Y are uncorrelated (meaning the correlation measure is 0), but rather X and Y are independent (which is stronger than uncorrelated).

If the null hypothesis were uncorrelation, X and Y could still be dependent; in that case, null distribution of the observed correlation would depend on the actual joint distribution of X and Y.

When X and Y are continuous and independent, the null distribution of the measures presented below does not depend on the actual joint distribution of (X, Y). This makes them useful in tests.

In stating the alternative hypothesis, bear in mind that the correlation measures do not reveal "association" (or "relation") in general, but only "monotonic" association (analogous to, but broader than, linear relation). To facilitate such statements, we introduce the following terminology.

Concordant: two observations (X_1, Y_1) and (X_2, Y_2) are concordant if $\frac{Y_1 - Y_2}{X_1 - X_2} > 0$.

Discordant: two observations (X_1, Y_1) and (X_2, Y_2) are discordant if $\frac{Y_1 - Y_2}{X_1 - X_2} < 0$.

With this terminology, the alternative hypothesis is pairs of observations tend to be concordant or discordant (two-sided), or concordant (one-sided), or discordant (one-sided).

3.1 Spearman's rho (ρ)

Simply use the ranks of X and Y, in place of the raw values, to calculate the Pearson correlation coef. Call it ρ.

If there are no ties, the expression of ρ can be somewhat simplified because we know terms like $\overline{R(X_i)}$ and $\sum_{i=1}^{n} R(X_i)^2$.

Test statistic: ρ.

Null distribution where there are no ties

Under H_0, the particular pairing between the observed X's and Y's is purely by chance. In other words, each of the $n!$ arrangements of the ranks of X_is paired with the ranks of the Y_is is equally likely. To find the null distribution of ρ, list all $n!$ possible rank orderings of Y, pair them with sorted X (with increasing ranks, $1, \ldots, n$), calculate ρ in these $n!$ cases, and tabulate the distribution.

Normal approximation

For large n, or many ties, use $E(\rho) = 0$ and $\text{var}(\rho) \approx \frac{1}{n-1}$.

3.2 Kendall's tau (τ)

Examine all the $\binom{n}{2}$ pairs of the observations. Define

$$\tau = \frac{N_c - N_d}{N_c + N_d + N_t}$$

where N_c is the number of concordant pairs, N_d is the number of discordant pairs, and N_t is the number of pairs in which $Y_1 = Y_2$ but $X_1 \neq X_2$ (that is, ties in Y). Tied pairs in which $X_1 = X_2$ are discarded.

If there are no ties, $N_t = 0$ and $N_c + N_d = \binom{n}{2} = n(n-1)/2$.

I have a complaint about τ: it treats X and Y differently when it comes to ties.

Test statistic: τ.

Null distribution when there are no ties

The exact null distribution can be found in a way analogous to what we did for ρ. In this case we could use $N_c - N_d$ instead of τ as the test statistic to save some computation (because τ is $N_c - N_d$ divided by $n(n-1)/2$, which does not affect the result).

Normal approximation

For large n or many ties, use $E(\tau) = 0$ and $\text{var}(\tau) \approx \frac{2(2n+5)}{9n(n-1)}$.

Remark Comparisons between Spearman's ρ and Kendall's τ:

1. τ has a simple and direct interpretation in terms of the probabilities of observing concordant and discordant pairs.

2. The normal approximation for τ is better than that for ρ. (The distribution of τ approaches normal rather rapidly; in fact, the approximation is considered quite good when used to find the quantiles of τ for $n \geq 8$, but not nearly as good when used to find the quantiles of ρ.)

3. In most cases, tests using either ρ or τ will reach the same conclusion.

3.3 Test for trend using ρ or τ

Given a time series $X_1, \ldots, X_n$, we can test for trend in X using the rank correlation ρ or τ between observation time and X. (The ranks of the times are simply $1, \ldots, n$.)

Remark Comparisons with the Cox and Stuart test:

1. When both methods are applicable, the test based on ρ or τ is generally more powerful than the Cox-Stuart test.
2. The test based on ρ or τ is not as widely applicable as the Cox-Stuart test. For example, this test does not apply to EX 3.5.3 (page 171).

4 Nonparametric methods for regression

Observations are $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Firstly and informally, when n is small, the following linear regression may be useful because it is more robust than the usual LS estimation.

1. For each pair of points (X_i, Y_i) and (X_j, Y_j) such that $i < j$ and $X_i \neq X_j$, compute the "two-point slope":

   $$S_{ij} = \frac{Y_j - Y_i}{X_j - X_i}$$

   Let N be the number of slopes computed. (If there are no X ties, N will be equal to $\binom{n}{2}$.)

2. Find the straight line that goes through the median point $(\tilde{X}, \tilde{Y})$ with a slope that is the median of the slopes $\{S_{ij}\}$.

4.1 Nonparametric test for the slope of a linear regression

Suppose the true regression equation is $E(Y) = \alpha + \beta X$. Now we don't need to obtain estimates of the coefficients; we just test on the slope:

$H_0 : \beta = \beta_0$

This test can be done by testing the significance of the Spearman's rank correlation ρ between X_i and the residual $U_i = Y_i - \beta_0 X_i$.

Note If the true regression function is $E(Y) = \alpha + \beta_0 X$, the residual is $E_i = Y_i - \alpha - \beta_0 X$. The unknown intercept α does not affect the test here because it would simply shift the U_i by the same amount for every i, an operation that does not affect the rank correlation we examine.

Note We could have used Kendall's τ instead of Spearman's ρ to do the test. The conclusion will be the same in most situations.

Example EX 1, page 336.

Exercise Compare this procedure with a parametric test (which requires the additional assumption of normal residuals).

4.2 Nonparametric confidence interval for the slope of a linear regression

After obtaining the "two-point slopes", a natural idea is to a "trimmed' range of these slopes as a confidence interval for the true regression slope. Suppose there are in total N two-point slopes (discarding point pairs where $X_1 = X_2$); order them as

$$S^{(1)} \leq S^{(2)} \leq \cdots \leq S^{(N)}$$

Then we'll take $[S^{(r)}, S^{(s)}]$, $1 \leq r < s \leq N$, as a confidence interval for the true slope. Naturally we'll cut off tails of equal length on both sides, hence $s = N - r + 1$. Therefore the CI is

$$\left[S^{(r)}, S^{(N-r+1)}\right]$$

The question is how to determine r.

Let $\beta_0 = S^{(r)}$. The hypotheses to be tested on are
$H_0 : \beta = \beta_0$
$H_a : \beta \neq \beta_0$

Then the test should barely reject on a $1 - \alpha$ level.

Calculate the residuals $U_i = Y_i - \beta_0 X_i$, then

$$S_{ij} = \frac{Y_i - Y_j}{X_i - X_j} = \beta_0 + \frac{U_i - U_j}{X_i - X_j}$$

i.e.

$$S_{ji} - \beta_0 = \frac{U_i - U_j}{X_i - X_j}$$

hence the sign of $\frac{U_i - U_j}{X_i - X_j}$ depends on whether S_{ij} is greater than or smaller than β_0).

When $\beta_0 = S^{(r)}$, the $N_c - N_d$ in a test for correlation between U_i and X_i is $(N - r) - r$. For this test to barely reject, we have

$$(N - r) - r = w_{1-\alpha/2}$$

where $w_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of $T = N_c - N_d$ when U_i and X_i are independent. From this we get

$$r = \frac{N - w_{1-\alpha/2}}{2}$$

4.3 Nonparametric monotonic regression

A usual goal of regression is to predict the Y in (X, Y) where X is known but Y is not. The nonparametric method presented here assumes the relation between X and Y is monotonic but need not be linear. It makes use of the fact that ranks preserves "orders" ("monotone") but ignores information regarding the actual distances (whether they are linear, quadratic, ...).

We first describe a method to predict the Y at a single X. After that we use the same idea to get a regression curve.

An estimate of $E(Y \mid X)$ at a point

The point has X value (x_0).

1. Obtain the ranks $R(X_i)$ of the Xs and $R(Y_i)$ of the Ys. Use average ranks in case of ties.
2. Obtain the LS linear regression on the ranks:
$$E[R(Y)] = a_2 + b_2 R(X)$$
3. Find the rank $R(x_0)$ among the observed Xs; use linear interpolation if needed, but do not extrapolate (stop if x_0 is outside the range of observed Xs).
4. Get the regressed Y rank: $R(y_0) = a_2 + b_2 R(x_0)$.
5. Get the y_0 value corresponding to the rank $R(y_0)$ among the observed Ys; use linear interpolation if needed.

Exercise The procedure is based on the fact that if two variables have a monotonic relationship, their ranks will have a linear relationship. Can you prove this fact? Can you demonstrate this fact empirically by constructing monotonic but nonlinear relations (using a computer and graphics)?

An estimate of the regression of Y on X

Observations are $(X_1, Y_1), \ldots, (X_n, Y_n)$.

1. For each X_i, $i = 1, \ldots, n$, obtain regressed $Y_{(i)}$ using the procedure above.
2. For each Y_i, $i = 1, \ldots, n$, obtain regressed $X_{(i)}$ using the procedure above. Note: use the inverse relation $R(X) = [R(Y) - a_2]/b_2$; do not obtain a new regression of X on Y.
3. Plot the points $(X_i, Y_{(i)})$ and $(X_{(i)}, Y_i)$, $i, \ldots, n$. Connect adjacent points. The whole curve should be monotonic.

Example EX 1, page 346.

4.4 Smoothing and local regression

Read the papers by Cleveland (1979, JASA) and Cleveland and Devlin (1988, JASA).

5 The one-sample or matched pairs case

The goal is to test whether the median or mean of X is zero (in the one-sample case) or whether X and Y have the same median or mean (in the matched pairs case). In the matched-pairs case, we actually have a bivariate random variable (X, Y) and we have a sample of it. The strategy is to take the difference, $D = X - Y$, in each pair, then it becomes a one-sample case.

In the signs test or median test, we used the counts of positive entries and a Binomial distribution.

Here we make an additional assumption: the distribution involved is symmetric. With this assumption, we can use more than just signs—we can use ranks.

Because of the assumption of symmetry, hypotheses can be stated in terms of either mean or median.

Wilcoxon signed ranks test

Take the differences (excluding differences that are 0):

$$D_i = X_i - Y_i, \quad i = 1, \ldots, n$$

Rank the D_is according to their absolute values, but let the rank carry the sign of D. For example, suppose the Ds are

$$2, 1, -3, 6, -4, 2, 5, 7, -8$$

Sort by absolute value:

$$1, 2, 2, -3, -4, 5, 6, 7, -8$$

Assign signed ranks R_i:

$$1, 2.5, 2.5, -4, -5, 6, 7, 8, -9$$

Test statistic when there are no ties and n is small

$$T = \sum R_i,$$

where $R_i > 0$ (i.e., sum of positive signed ranks).

Under the null hypothesis that $E(D) = 0$, we can imagine that T should be roughly half of $1 + \cdots + n$. How can we find the exact distribution of T?

Because $E(D) = 0$, any rank has equal probability to come from a positive or negative observation. If we list the ranks and assign signs to them, they independently receive '+' or '-' with equal probability. There are 2^n ways to assign signs; based on that we can tabulate the distribution of T.

Test statistic when there are many ties or n is large

$$T = \frac{\sum R_i}{\sqrt{\sum R_i^2}}$$

using all signed ranks (both positive ones and negative ones).

Take the a_ith rank, that is, $|R_i| = a_i$, then under H_0 we have $P(R_i = a_i) = P(R_i = -a_i) = 1/2$, hence

$$E(R_i) = 0, \quad \text{var}(R_i) = a_i^2$$

Therefore $E(\sum R_i) = 0$. Using the independence between the R_is, we see $\text{var}(\sum R_i) = \sum R_i^2$ which is $1^2 + \cdots + n^2$ when there are no ties.

Invoking CLT, T is approx standard normal.

6 ANOVA using ranks

Parametric ANOVA are based on normality assumptions. In this section we briefly discuss two two-way ANOVA approaches that do not assume normality—they use ranks instead of the raw data.

6.1 Randomized complete block design; the Friedman test

		Treatment		
Block	1	2	$\cdots$	k
1	X_{11}	X_{12}	$\cdots$	X_{1k}
2	X_{21}	X_{22}	$\cdots$	X_{2k}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
b	X_{b1}	X_{b2}	$\cdots$	X_{bk}

b blocks; k treatments. Each block contains k "experimental units", which are randomly assigned to k treatments.

The goal is to test whether there is a "treatment effect", that is, the different treatments lead to different results, in other words, the treatments "make a difference", have an "effect". (Existence of treatment effect does not mean that different treatments have different effects. Rather, it means the k treatment are not all the same—at least two show difference.)

The blocks are to "control" some factor, so that in each block the treatment each experimental unit receives is the main factor that is different among them. Consequently, if the units in a block show difference, the difference is likely caused by the difference in the treatments. This ability to infer "causality" is provided by that the assignment of treatments to units (or units to treatments) is random.

This arrangement is called randomized complete block design. It's "complete" because all treatments are represented in each block. If there are more treatments than experimental units in a block, then only some of the treatments will appear in one block; that would be an "incomplete" design.

Example Page 368: psychology, home economics, environmental engineering

Test statistic:

Rank the X's within each block (row); call the ranks $R(X_{ij})$. Get the column totals:

$$R_j = \sum_{i=1}^{b} R_{X_{ij}}$$

Let

$$T = \frac{6}{k}\left(\frac{b(k+1)}{2}\right)^{-1} \sum_{j=1}^{k}\left(R_j - \frac{b(k+1)}{2}\right)^2$$

Note Under the null hypothesis, $E\big(R_{(}X_{ij})\big) = \frac{1+k}{2}$, hence $E(R_j) =$

$\frac{b(k+1)}{2}$. Compare T_1 to the χ^2 tests we learned before.

Note In the matched-pairs case, we tested whether $E(X) = E(Y)$. Here, we test whether the means of the treatment levels are all equal. The current situation is a generalization of the matched-pairs case. In the latter, the different bivariate observations are "blocks"; each block contains two "experimental units', one for each "treatment" (X or Y).

Null distribution

Approximately,

$$T \sim \chi^2_{k-1}$$

Exact null distribution

Under the null hypothesis, each ranking with any block is equally likely. There are $k!$ possible arrangements of ranks $R(X_{ij})$ within a block and, therefore, $(k!)^b$ possible arrangements of ranks in the entire array of b blocks. Therefore, the distribution of T_1 can be found by listing all possible arrangements of ranks and computing T_1 for each arrangement.

Example EX 1, page 371.

Improvement

The χ^2 distribution approximation for T_1 is sometimes poor. A preferred alternative test statistic is calculated by the usual parametric ANOVA, only using the ranks instead of the raw data. The resultant test statistic turns out to be a function of T_1; let's call it T_2. T_2 has an approximate F distribution. The distribution approximation is better than the χ^2 approximation for T_1.

6.2 Balanced incomplete block design; Durbin test

Example EX 1, page 390.

Features of the design:
(1) There are b blocks;
(2) Each block has the same number (k) of experimental units;
(3) There are t treatments; $t > k$.
(4) Each treatment appears in r blocks; $r < b$. (5) Every treatment pair appear in the same number of blocks. For example, treatment-pair 1 and 3 appears only in block 3. Similarly, every other treatment-pair also appears only once.

These features make it a balanced incomplete block design.

Test statistic

Rank the units in each block (i.e. row).
Get the total rank of each treatment:

$$R_j = \sum_{i=1}^{b} R(X_{ij}), \qquad j = 1, \ldots, t$$

(If j indicates treatment j, then in a block it's possible that no experimental unit is subject to treatment j. In that case consider $R(X_{ij})$ to be 0.)

Let

$$T_1 = \frac{6(t-1)}{t(k-1)} \left(\frac{r(k+1)}{2} \right)^{-1} \sum_{j=1}^{t} \left(R_j - \frac{r(k+1)}{2} \right)^2$$

Note that $E\big(R(X_{ij})\big) = \frac{k+1}{2}$, hence $E(R_j) = \frac{r(k+1)}{2}$.

T_1 has approximately a χ^2_{t-1} distribution.

Exact null distribution

Under the null hypothesis, each ranking with any block is equally likely. There are $k!$ possible arrangements of ranks $R(X_{ij})$ within a block and, therefore, $(k!)^b$ possible arrangements of ranks in the entire array of b blocks. Therefore, the distribution of T_1 can be found by listing all possible arrangements of ranks and computing T_1 for each arrangement.

Improvement

Similar to the Friedman's test, compute the usual parametric ANOVA test statistics, but replace the raw values by the ranks. The resultant test statistic turns out to be a function of T_1; let's call it T_2. T_2 has an approximate F distribution. The distribution approximation is better than the χ^2 approximation for T_1.

7 Randomization tests

Also called "permutation tests".

Use the original data, not ranks. Key point is "exchangeability". Often intuitive; available for any statistic. Computationally intensive.

It's best to learn this idea by examples.

Example EX 1, page 410.

Example EX 2, page 413.

8 Some connections between parametric and rank-based nonparametric methods

Many rank-based nonparametric methods come directly from parametric method: just use ranks in the formulas of the parametric methods.

The most obvious examples we have seen include (1) Spearman's rank correlation; (2) monotonic regression.

Additional examples: (1) Mann-Whitney test vs two-sample t test; (2) Wilcoxon signed ranks test vs one-sample t test. See page 417–419.